MULTIMEDIA UNIVERSITY®

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2016/2017

HPB3011 –Bioinformatics Algorithms II

(All sections / Groups)

10 OCT 2016
2:30 – 4:30 PM
[2 hours]

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 4 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal marks (10).
3. Write all your answers in the Answer Booklet provided.
4. You may use the calculator.

Question 1

a) What are the two main types of microarray gene chips? [2 marks]

b) What are the types of algorithm used for hierarchical clustering? Briefly explain the distance used in hierarchical clustering for clusters. Explain single linkage clustering using any example. [5 marks]

c) What is the step commonly used to reduce systemic variations in DNA microarray data analysis? Name two methods used in the step and explain briefly. [3 marks]

Question 2

a) What are the criteria for deciding if a sequence is a CpG island? Why are such CpG sequences called "islands"? Why are they biologically interesting? [3 marks]

b) In a Support Vector Machine (SVM), what are the "support vectors"? How is the optimal dividing line or plane found - what function is involved? How could you use SVMs to separate more than 2 classes (e.g. helix, sheet, coil or turn)? If your 2-dimensional data points cannot be separated in 2 dimensions, what could you do to still separate them with a SVM? [4 marks]

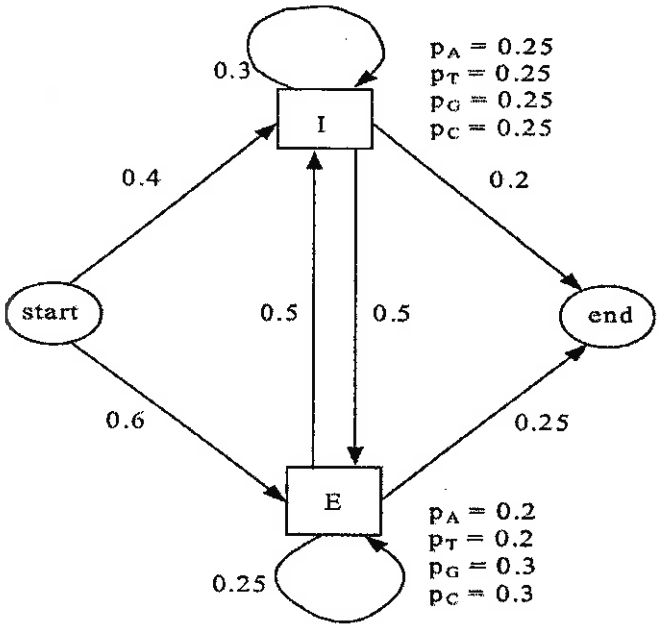
c) Given the following probabilities, use the Bayes approach to find the probability of having diabetes if the test is positive - P(Diabetes | test+). [3 marks]

	Diabetes	NO Diabetes
Test +	0.8	0.01
Test -	0.2	0.99

Given that the overall probability of diabetes is 0.01.

Continued...

Question 3



a) Find the probability of the DNA segment "TCA" predicted by the hidden Markov model with the state-transition diagram given in the figure above, using the Forward Algorithm [5 marks]

b) Define ANN - Artificial Neural Networks. Draw a simple diagram of a 2-layer ANN (input layer, hidden layer, output). [3 marks]

c) Explain the steps in supervised learning to train the ANN for 2 different input signals. [2 marks]

Question 4

a) You have discovered a new protein for which you have gotten the amino acid sequence. It has 80% sequence identity to another protein, which is known to be an enzyme. Your protein, however, appears to recognize a different class of substrates. Is this likely to be convergent or divergent evolution? Explain.[4 marks]

b) What are the basic concepts behind Chou-Fasman method? Describe the steps in the CF method to determine which amino acid sequences are likely to be α–helices, β–sheets, describe the criteria for decisions. [4 marks]

c) What is Protein structure prediction? Explain 4 importance of protein structure prediction? [2 marks]

Continued...

Question 5

a) What are the steps in the protein threading method? [4 marks]

b) Discuss the function of RNA chaperones. How are they relate to the functions of RNA molecules? [4 marks]

c) What is protein folding? Name one disease caused by misfolding of protein. [2 marks]

END OF PAPER